Overall inefficiency and cycles in non-ordered screening under capacity constraints and standardization

By

Sergey Kokovin[a], **Babu Nahata**[b], **and Evgeny Zhelobodko**[a]

[a]Department of Economics, Novosibirsk State University, Novosibirsk 630090, Russia,

e-mail: kokovin@math.nsc.ru, ezhel@ieie.nsc.ru

[b]Department of Economics University of Louisville, Louisville, Kentucky 40292, USA.

e-mail: nahata@louisville.edu.

Abstract. We study discrete-type screening without Spence-Mirrlees condition. Under non-separable and non-concave cost, *all* packages can be distorted in equilibrium, even when only the participation constraints are active. This and other paradoxical effects, shown by examples, are caused by some kind of *envy-cycles* among agents. Theorem 1 proves that such effects are precluded under separable or concave cost, thus justifying the applicability of the standard screening model, which appears doubtful under more general costs.

Keywords: multidimensional screening, socially inefficient price discrimination, monopolistic standardization, graph cycles.

JEL Codes: D42, D82, L10, L12, L40.

1 Introduction

There is a vast literature on screening. Recently, more attention has been focused on situations where agents' utilities are non-ordered (without requiring single-crossing condition) or when goods are multidimensional (see, the most recent and comprehensive review by Rochet and Stole (2007)). Yet, this rich and growing literature has paid almost no attention to screening problem when the seller faces some capacity constraints, which make costs non-separable with respect to agents and non-concave. Such problems are quite common in real life, for at least two reasons. First, the *upper limit* on total production makes cost function increasing steeply to infinity, thus making it non-concave and non-separable. For instance, an airline can take only a limited number of passengers in a flight and this makes costs dependent on total number of flying passengers. Second, *standardization*, by limiting the number of sizes results in economies. For instance, offering only three classes of services (e.g., economy, business and first) may be more economical for an airline than many distinct classes. One can find many other examples in real life where either capacity constraint or standardization or both become important considerations in solving the screening problem.

Our examples demonstrate the theoretical complexities in dealing with these important situations, previously ignored in the screening literature. The strong motivation to consider such examples stems from the fact that under non-separable costs the standard methodology used in screening literature and the related well established theoretical results become questionable. In particular, cycles which are precluded or reducible under standard (separable) costs *worth study* under non-separable costs.[1] In this case, our examples show that cycles have interesting theoretical and practical consequences, most importantly on efficiency. The well-known "efficiency-at-the-top" result typical under all standard situations, may now be wrong, but, more surprisingly, even *all* packages may be distorted! This paradoxical result inspires us to revisit Dupuit (1849), who observed that, "... third-class... has open carriages with wooden benches... What the company is trying to do is prevent the passengers who can pay the second-class fare from travelling third-class; it hits the poor... to frighten the rich. And... having proved almost cruel to third-class and mean to second-class ones, become lavish in dealing with first-class passengers. ... they give the rich what is superfluous." [J. Dupuit, (1849): On Tolls and Transport Charges, quoted in Tirole (1994), p.150.]

The natural reaction of any economist, familiar with the screening literature, would be that Dupuit's assertion lacks theoretical support because overall inefficiency is impossible. But, our examples generally do support Dupuit's idea when costs are non-separable, and cost of passengers travelling are really non-separable (however, in this particular example overall distortion seems unrealistic for other reasons explained after Examples).

Another surprisingly result for the economists familiar with the screening literature is:

[1] Cycles or curcuits that we discuss, relate to some graph structure generated by the active (becoming equalities) incentive-comptaibility constraints at equilibrium, as explained in Section 2.

There *can* be an equilibrium where all packages are distorted, even though *all* incentive constraints are non-active!

The third unusual effect stemming from non-separable cost is a new kind of price discrimination — not only the standard discrimination *among* different types of consumers, but also discrimination *within* a homogeneous group of agents. Such examples calls for adjusting not only the screening model itself, but also the notion of its equilibrium to incorporate new situations analyzed in this paper.

2 Model and definitions

Consider the standard discrete-type screening problem. A monopolist offers a product or service using a menu of several discrete *packages* of different quantities or qualities at some fixed tariffs on take-it-or-leave-it basis and consumers self-select. Consumer types are indexed by $i \in I^n = \{1, ..., n\}$; and $m_i > 0$ is the frequency of type i. The consumption-outlay bundles are (x_i, t_i), where $x_i \in X$ denotes consumer i's consumption and $X \subset \mathbb{R}^l$ is a consumption set (in our examples mostly $X = \mathbb{R}$, which is more challenging than the multidimensional paradoxes, also constructed by us, but not included here). Tariff t_i is monetary transfer from consumer i to the firm. For simplicity, we assume quasi-concave utility functions $u_i(x_i, t_i) = v_i(x_i) + t_i$, where v_i is "monetary valuation" of commodity, normalized so that $v_i(0) = 0$, with no other assumptions. The principal knows the characteristics of types, but is unable to discriminate personally.

The producer's cost function $C(m, x)$ is of general form, which may have a special *aggregate* form $C(m, x) = c(\sum_i m_i x_i)$, where function $c : \mathbb{R} \to \mathbb{R}$. Another important form can be a generalized *separable* cost function $C(m, x) = f_0 + \sum_i m_i c(x_i)$, where f_0 stands for some fixed cost.

We start with the standard assumption that the producer, for whatever reasons, designs exactly n packages assigned to consumers by planning an *assignment*, $(x, t) = \{(x_i, t_i)\}_{i=1...n}$ and assigns only one package to each type. We also use the standard so-called "friendly-agent" assumption (i.e., an agent prefers the principal's choice, when having multiple equivalent choices). Later on both of these assumptions are questioned. Such optimization problem of the seller is labelled here as standard assignment-optimization program or SAOP (though the cost function is not standard), and formulated as follows.

$$\begin{aligned}
\pi(x,t) := & \quad \sum_{i=1}^{n} m_i t_i - C(m, x_1, ..., x_n) \to \max_{(x,t)\in(X^n,\mathbb{R}^n)}, \textbf{s.t.} & (1)\\
\forall i \in I^n \Rightarrow v_i(x_i) & - \quad t_i \geq v_i(x_k) - t_k \quad \forall k \in I^n \cup \{0\} \setminus \{i\}, & (2)\\
(x_0, t_0) := & \quad (0,0).
\end{aligned}$$

A solution $(\bar{x}, \bar{t})$ to this problem is the (standard) *equilibrium*.

Package labelled #0 here means no participation by the agent, and by including this dummy agent, all incentive-compatibility (IC) and participation constraints can be represented by equations (2).

An equilibrium can be characterized by its set of active constraints. Any active constraint (i, k) (i.e., constraint becoming an quality) is interpreted as "almost-envy" from i to k and treated as the directed arc $(i \to k)$ of an *almost-envy graph*, where agents or packages are represented by nodes, and constraints are shown as directed arcs, while (0) is the "sink." (Directed) *cycle* or circuit is a closed path $i \to k \to ... \to i$.

In a package-menu $(\bar{x}, \bar{t})$, we call one of its packages $(\bar{x}_i, \bar{t}_i)$ as socially *efficient* when it maximizes the sum of this consumer's and producer's surpluses (joint welfare), all other packages being fixed, i.e., $\bar{x}_i \in \arg\max_{z_i \in X} \left(v_i(z_i) - C(m, \bar{x}_1, ..., z_i, ..., \bar{x}_n)\right)$. Otherwise the package is *distorted* (over-sized or under-sized). The *overall-efficient* package-menu $(\bar{x}, \bar{t})$ maximizes total welfare, it does not contain any distorted packages and the *overall-inefficient* is the menu that does not include any undistorted (efficient) package.

Now we step aside from the above standard setting and consider an *extended screening model*, which is motivated by examples below.

A package *menu* is a bundle of quantity-stock-tariff triples $\{(q_k, s_k, t_k)\}_{k=1,...,K} \in X^n \times \mathbb{R}^{2n}$. It means that the principal chooses $K \geq n$ to design as many as K types of packages and produces a stock $s_k \geq 0$ of each package, without a priori deciding which package goes to which agent (though $x_i = q_i$ is typical). When agents come in a random sequence, some stock s_k may run out, so the remaining agents may be forced to choose from the reduced menu. Such distinction between "menu" and "assignment" becomes important when cycles appear or/and when the principal decides to *partition* some consumer type into subgroups taking different packages.

These new pricing tools become quite important under capacity constraints, unlike the standard setting. We are not outlining in detail the new relevant model and its equilibrium notion (it is too cumbersome). Instead, our modest goal is to give some hints and ideas through examples.

3 Overall inefficiency, cycles and type-splitting under non-concave costs

Three examples below show three types of cycles caused by three different peculiarities, namely by type-splitting, non-convexity and insufficient partitioning, in addition to capacity constraint present in all the three.

[FIGURE 1]

Generic inefficient cycle under convex cost.

Example 1. *Generic cycle and overall inefficiency.*

Consider a homogeneous good. Cost of production is normalized to zero, but the total feasible production is limited to 142.5 units by capacity constraint. Thus, $C(x_\Sigma) = 0$ for $x_\Sigma \leq 142.5$, $C(x_\Sigma) = \infty$ for $x_\Sigma > 142.5$.[2]

Group numbers (m_1, m_2, m_3) and valuations of the three types are described by the active indifference curves in Fig.1. (here $z \in R$):

$m_1 = 1$	$v_1[z] = \min\{9z,\ 1.7 + 0.5z,\ 28 - 3z\}$	long-dashed curve
$m_2 = 10$	$v_2[z] = \min\{0.745z, 5 + \frac{1}{24}z, 6.12 - \frac{1}{9}z, 29 - 3.1z\}$	dotted curve
$m_3 = 9$	$v_3[z] = \min\{4z, 2.2 + 0.38z, 5.5 - \frac{1}{15}z, 56.2 - 6.4z\}$	short-dashed curve

[2] It is easy to generalize this example to positive costs and strictly increasing utilities by adding same linear component $7x$ to all functions.

If the seller cannot *partition* these types into subgroups (see comment after the Example), then the optimal assignment based on SAOP is: $x = (x_1, x_2, x_3) = (17/35,\ 36/5, 705/91) = (0.486, 7.2,\ 7.747)$, and $t = (t_1, t_2, t_3) = (68/35,\ 53/10,\ 907/182) = (1.943,\ 5.3,\ 4.984)$. The three points are represented by thick dot, hollow dot, and rectangle respectively. Solution is obtained by directly solving SAOP (computer program is available from the authors, together with calculations for this and other examples). Related menu is: $(q, s, t) = ((17/35, 1, 68/35), (36/5, 10, 53/10), (705/91, 9, 907/182))$. The total quantity sold is $64706/455 = 142.211$ units (about 0.3 capacity is wasted), and the total profit is $\pi = 90813/910 = 99.795$. The almost-envy-graph (shown in Fig.1) is $[\#1 \rightarrow \#2 \rightarrow \#3 \rightarrow \#1,\ \ \#1 \rightarrow 0,\ \ \#2 \rightarrow 0,\ \#3 \rightarrow 0]$, and it contains the *envy-cycle* among agents $\#1, \#2$ and $\#3$.

It is this cycle that results in *overall distortion* including *inefficiency at the top* (of the graph). Indeed, nobody takes the socially-efficient quantities, which are: $x_1^*= 7.5$, $x_2^*= 7.3$, $x_3^*= 7.4$, respectively. Graphically, this overall inefficiency means that equilibrium points are not the peaks of the three active indifference curves.

Looking at this example in terms of extended screening model discussed above, note that a better *partitioning* of agents can eliminate most of distortion as well as cycle, thus increasing both profit and social welfare. Indeed, a more profitable pricing strategy is to split type-3 consumers into groups, but bunch type-1 and type-2 together, by designing a (non-optimal) contract $(\hat{q}, s, \hat{t}) = ((\frac{94}{91}, 1, \frac{5898}{2275} - \frac{3}{15}\varepsilon), (\frac{36}{5}, 11, \frac{53}{10} - \frac{1}{15}\varepsilon), (\frac{705}{91} + \varepsilon, 8, \frac{907}{182} - \frac{3}{15}\varepsilon))$ with small $\varepsilon > 0$ to get strict incentive compatibility among types (not groups). It results in a feasible assignment that brings profit of $\pi \approx 100.76$, which is higher than from the initial contract (q, s, t).

The SAOP screening method is inadequate here and also in similar situations for two reasons. First, non-trivial partitioning of types can be more profitable than usual SAOP method "one package for one type." Paradoxically, here price discrimination is practiced to discriminate even between *identical* agents, and this partitioning increases profit and total welfare!

Second, consider the "friendly-agent" assumption, so common in screening and optimal contracts. Usually it is justified, because the principal can approximately make the optimal plan strictly incentive compatible by reducing tariffs, using small $\varepsilon-$*rewards for friendship.*

But such modifications of tariffs can not solve the problem of motivating friendship under cycle, as one can see in the example. So, implementing the initial plan (q,t) becomes problematic, unless modifying quantities also, like in our second plan $(\hat{q},\hat{t})$, to implement the menu in dynamic game. This plan also uses the *stock* tool of pricing: when one package $(\hat{q}_1,\hat{t}_1)=(\frac{94}{91},\frac{5898}{2275}-\frac{3}{15}\varepsilon)$ is bought, nobody can further buy such a package. Such tool can enhance profit relative to SAOP "optimal" menu (so, SAOP becomes inadequate).

To comprehend dynamic implementation and its importance, consider the initial menu (q,s,t). What happens if 9 agents of type-2 take all 9 packages assigned for type-3 before agents of type-3 arrive? Then nine type-3 agents go away, causing a substantial profit loss relative to the SAOP profit $\pi=99.795$ (see also the discussion of implementation after Example 3). Thus SAOP-based profit (1) is not guaranteed, (2) can be improved. ||

It is worth noting that Example 1 is generic in weak sense that any small disturbances in three typical valuations (but not in 20 individual valuations!) and in costs cannot eliminate the cycle and inefficiency resulting from using SAOP. However, if standardization (not modeled explicitly) forces the principal to design only three packages, one for each group, then this example can be generic in all senses.[3] Another observation is that here inefficiency results from an unprofitable partitioning and cycle at the top, but paradoxically, a better partitioning does not require increasing number of packages. Example 2 below differs in this respect, and also in the nature and causes of overall inefficiency. There, overall inefficiency results from the requirement of standardization to integer quantities, and from the capacity constraint.

EXAMPLE 2. *Shadow cycle (binding-constraints cycle) and overall inefficiency, in spite of no active IC constraints.*

Suppose that only five discrete quantities of total production of a homogeneous good are economically feasible: $x_{sum}=0$, 1, 2, 3, 4 or 5, with per-package costs normalized to zero as $C(0)=C(1)=C(2)=C(3)=C(4)=C(5)=0$ at discrete points and prohibitively high costs otherwise. Total production is limited as $x_{sum}\leq 26$.

[3] For example, a developer of a piece of land may be forced to offer only three standards: big lots for rich people, medium-size for middle class, and small lots for poor, because all prefer to be surrounded by the same class. Then the assumption one-package-for-one-type is an *exogenous* restriction.

There are three ($m_1 = 3$) agents of type #1 and $m_2 = 4$ of type #2. Their valuations are $v_1(z) = \min[5z - 5,\ 3 + z]$, $v_2(z) = \min[4.2z - 4.2,\ 2.6z - 2,\ 3.8 + z,\ 7.72 + 0.2z]$. They also can be defined point-wise at 5 admissible quantities (1, 2, 3, 4, 5) as:

3 of #1	$v_1(1) = 0$	$v_1(2) = 5.0$	$v_1(3) = 6.0$	$v_1(4) = 7.0$	$v_1(5) = 8.0$
4 of #2	$v_2(1) = 0$	$v_2(2) = 4.2$	$v_2(3) = 6.8$	$v_2(4) = 7.8$	$v_2(5) = 7.82$

These points are squares and circles in Fig. 2, interpolated by lines to become $v_2(.)$ (dashed) and $v_1(.)$ (dotted, with circles).[4]

[FIGURE 2]

Generic inefficient shadow-cycle under discrete cost.

Direct calculation shows that the only optimal assignment based on SAOP (shaded points in Fig. 2) is: $\{(\bar{x}_1, \bar{t}_1), (\bar{x}_2, \bar{t}_2)\} = \{(2,\ 5.0), (4, 7.8)\}$ with total production $\bar{x}_{sum} = 22$ and profit $\pi(\bar{x}, \bar{t}) = 46.2$. The alternative assignments are shown by hollow points, but it is easy to check that they cannot be included into the optimal plan due to IC constraints combined with the upper bound of 26 on total production. Assignment $(\bar{x}, \bar{t})$ is strictly incentive compatible, so it is free of "almost-envy." Yet, both packages are inefficient in the sense that Pareto-improvement could be possible. There are two technologically feasible incentive-incompatible non-standard menus more efficient than $(\bar{x}, \bar{t})$: $\{(\tilde{x}_1, \tilde{s}_1, \tilde{t}_1), (\tilde{x}_{2'}, \tilde{s}_{2'}, \tilde{t}_{2'})\} = \{(2, 3,\ 5.0),\ (5, 4,\ 7.82)\}$, $\{(\hat{x}_{1'}, \hat{s}_{1'}, \hat{t}_{1'}), (\hat{x}_2, \hat{s}_2, \hat{t}_2)\} = \{(3, 3,\ 6.0), (4, 4,\ 7.8)\}$.

So, surprisingly, in spite of the menu $(\bar{x}, \bar{t})$ being envy-free, the usual efficiency-at-the-top property does not hold for both packages. It is because the non-active constraints are binding and constitute a *shadow cycle,* shown in the right panel.

SAOP with naive (standard) partitioning here results in loss of profit as in Example 1. Indeed, if the monopolist is able to use three packages instead of two, and better partitioning, then, unlike Example 3 below, such *better partitioning and stocks* can resolve the problems with cycle. In particular, menu $\{(\hat{x}_1, \hat{s}_1, \hat{t}_1), (\hat{x}_2, \hat{s}_2, \hat{t}_2), (\hat{x}_{1'}, \hat{s}_{1'}, \hat{t}_{1'})\} = \{(2, 3,\ 5.0), (4, 4,\ 7.8), (5, 1,\ 8.0)\}$ is incentive-compatible, implementable under rationing

[4] Here, valuations are continous and costs are discrete, otherwise one can make costs continuous, making valuations positive only at discrete points. Some discreteness or non-convexity and different numbers $m_1 \neq m_2$ of consumers are essential for constructing such an example. Discreteness is natural when feasible are only 1 box, 2 boxes, 3 boxes, etc.

by stock, and more profitable than $(\bar{x}, \bar{t})$. This stock means *rationing,* it makes additional profit by forcing some group of agents split into two sub-groups buying different packages, so discriminating *within* one type.

However, this pricing tool fails when it is too-costly for the seller to use more than $K = 2$ packages (this standardization requirement is modelled as restriction $K \leq 2$). Then the inefficient $(\bar{x}, \bar{t})$ remains the only solution. Note that this example is non-degenerate, in the sense that slight perturbations of data do not change anything. Small perturbations of individual consumers can make five consumer types out of two types in this example, but this also does not matter, if standardization does restrict the number of packages K. ||

In summary, even when all packages are envy free (which is often supposed to result in overall efficiency) they can produce a shadow-cycle and overall *inefficiency*, including inefficiency at the top! Again, this effect is due to insufficient partitioning and rather specific costs, connected with *standardization.*

EXAMPLE 3. *Non-reducible cycle at the top resulting from non-concave costs and valuation.*

Instead of inefficiency, this example shows that non-reducible cycle can appear because of non-concave cost, without additional specific reasons such as standardization or insufficient partitioning of population. This cycle can bring profit loss like in Example 1 (that show inadequacy of SAOP).

[FIGURE 3]

Non-reducible cycle for locally-similar agents and convex cost.

Suppose the aggregate cost function is of the type: $C(x_{sum}) = \max\{0,\ 3(x_{sum} - 6.5)^3\}$. Thus, any production less than 6.5 is costless, but it follows from below data that total production $x_{sum} := \sum_i x_i > 7.75$ is prohibitively costly. There are five buyers of 5 types with the following valuations:[5]

[5] As in other examples, one can add a linear component Ax to the cost function and to all valuations to keep the essence of the example intact without requiring satiation for agents.

$v_1(x_1) = 4\,x_1 - 3.325\,x_1^2$ (solid parabola), $v_2(x_2) = 1.32\,x_2 - 131/360\ \ x_2^2$ (solid parabola), $v_3(x_3\,) = \min\{5\,x_3\,;\ 1 + 0.01x_3\ ;\ 12.8 - 4\,x_3\,\}$ (thin solid curve),[6]

$v_4(x_4\,) = \min\{5\,x_4\,;\ 1 + 0.01x_4\ ;\ 4 - 2\,x_4\,\}$ (dashed curve),

$v_5(x_5) := \min\{2\ x_5\,;\ 1 + 0.01x_5\,;\ 12.8 - 4\,x_5\}$ (dotted curve).

An optimal menu with total output $x_{sum} = 98/15 \approx 6.533333$ is: $(\bar{q}, \bar{s}, \bar{t}) = ((0.6, 1, 1.203), (1.8, 1, 1.197), (0.25, 1, 1.0025), (1.0, 1, 1.01), (173\ /60, 1, 6173/6000))$ with related assignment shown in Figure 3: $(\bar{x}_1, \bar{t}_1) = (0.6, 1.203)$, $(\bar{x}_2, \bar{t}_2) = (1.8, 1.197)$, $(\bar{x}_3, \bar{t}_3) = (0.25, 1.0025)$, $(\bar{x}_4, \bar{t}_4) = (1.0, 1.01)$, $\bar{x}_5 = 173\ /60 \approx 2.883\,33$, $\bar{t}_5 = 6173/6000 \approx 1.0288333$, resulting in a cycled envy graph: $\#i \to 0 \forall i$, $\#4 \leftarrow \#5 \leftarrow \#3 \leftrightarrows \#4$.

To see that this assignment is *optimal* in SAOP, note that the entire consumer surplus goes to the principal and total welfare is maximized. Indeed, the derivatives of all valuations and costs are the same: $v_i'(\bar{x}_i) = C'(98/15) = 0.01$, which suffices for optimum in optimization of $v_i(x_i) - C(x_i)$ which is joint welfare of a couple: consumer and producer. One can also verify by trial and error that, except for permuting agents #3, #4 and #5 around the envy-cycle (that gives same profit), or for very small changes in x_3, x_4, and x_5, this $(\bar{x}, \bar{t})$ is essentially a unique first-best solution.

Here some sort of type-splitting is inevitable in all solutions of SAOP, i.e., different packages should be offered to locally-similar agents #3 and #4. The same goes for couple #4, #5. This similarity produces the envy-cycle (in Fig.3, the dash arrow shows these ties for only one example version, while solid arrows relate to both) which is not generic.

Here cycle in itself does not bring inefficiency or direct loss of profit. However, indirect loss stems from cycle's influence on *implementation* of the principal's optimal menu, like in Example 1. What happens if agents #3, and #4 take packages #5 and #4 respectively, before agent #5 comes? Then this type-5 agent goes away unserved, causing a profit loss of \$1.0288. This non-implementation problem can be resolved by making packages #3, and #5 slightly cheaper than #4. But if we slightly modify $v_3(.)$ as $\tilde{v}_3(x_3\,) = \min\{(0.95 + (x-1)^2); 5\,x_3\,;\ 1 + 0.01x_3\ ;\ 12.8 - 4\,x_3\,\}$, then optimum remains the same but cycle and non-implementation problem *cannot be resolved* by any means, as one can check.

[6] The cavity in $v_3(.)$ on the graph relates to some curve modification: $\tilde{v}_3(x_3\,) = \min\{v_3(x_3\,); 0.95 + (x-1)^2\}$ explained later on. Optimum is the same for both versions.

We see that cycles undermine the traditional "friendly agent" and "no-type-splitting" assumptions, the basic assumptions of SAOP model, which becomes inadequate here.

To conclude the discussion of examples, we can explain why Dupuit's conjecture from Introduction seems wrong. To generate overall inefficiency, the railroad example should have one of the specific features similar to those in our three examples, and exhibit either almost-envy cycle, or binding-envy cycle. Instead, under single-crossing condition (SMC) absence of cycles and efficiency at the top are guaranteed, standardly, whereas SMC seems very realistic for passengers. Indeed, high-income travellers not only wish to pay more for comfort, but they also wish to pay more for an additional unit of comfort.

4 No paradoxes under separable or concave cost

In contrast to previous section, we now focus on situations where the standard SAOP model does work well. The standard literature on screening theory, assuming linear cost, can (and does) ignore the above mentioned methodological hardships: non-reducible cycles, non-implementation and type-splitting (partitioning). The only formal justification for such simplified approach that we know, relates to proving reducibility of cycles in Guesnerie and Seade (1982), generalized further in Brito et al. (1990), and in Andersson (2005) under linear cost. Theorem 1 below extends this reducibility result to separable or concave costs, and adds no type-splitting claim. Thus, our Theorem provides a justification for the use of traditional SAOP assumptions (namely, friendly-agent and no-splitting) in usual situations as well as in some new situations, thus expanding the area of SAOP application up to the limits of its use (shown by the examples).

Theorem 1 needs two notions. When costs are separable with a fixed-cost component $f_0 \geq 0$ ($C(m,x) = f_0 + \sum_{i=1}^{n} m_i c_i(x_i)$), the artificial *per-package profit* function $\bar{\pi}_i$ is defined as:

$$\bar{\pi}_i(x,t) := t_i - c_0 - c_i(x_i), \;\; \text{where} \;\; c_0 := f_0 / \sum_{i=1}^{n} m_i \ .$$

For an arbitrary cost function $C(.)$, a similar notion is introduced by linearizing C at some given point $\bar{x}$:[7]

$$\bar{\pi}_i(x,t) \equiv \bar{\pi}_{i\bar{x}}(x,t) := t_i - c_0 - \bar{c}_i x_i \ , \ \text{ where } \bar{c}_i := \frac{1}{m_i}\frac{\partial C(m,\bar{x})}{\partial x_i}, \ \ c_0 := \frac{C(m,\bar{x}) - \sum_{i=1}^n m_i \bar{c}_i \bar{x}_i}{\sum_{i=1}^n m_i}.$$

This notion enables eliminating cycles from the package-graph without any loss in profit, by using the following bunching procedure like in Guesnerie and Seade (1982).

Bunching procedure. In a feasible situation $(\bar{x},\bar{t})$, when an agent j almost-envies somebody i $(j \to i)$ then we can replace j-th assignment $(\bar{x}_j,\bar{t}_j)$ with a new (envied) package $(\hat{x}_j,\hat{t}_j) = (\bar{x}_i,\bar{t}_i)$, keeping other components unchanged. The complete new assignment $(\hat{x},\hat{t})$:=$((\bar{x}_1,\bar{t}_1)$, ..., $(\bar{x}_{j-1},\bar{t}_{j-1}), (\bar{x}_i,\bar{t}_i), (\bar{x}_{j+1},\bar{t}_{j+1})$, ..., $(\bar{x}_n,\bar{t}_n))$ remains incentive compatible, because no new packages appears in the menu and all other agents remain unaffected, except for j. This agent j has exactly the same payoff as before: $v_j(\hat{x}_j) - \hat{t}_j = v_j(\bar{x}_j) - \bar{t}_j$, because of the envy arc $(j \to i)$, so, incentive compatibility holds: $v_j(\hat{x}_j) - \hat{t}_j \geq v_j(\bar{x}_k) - \bar{t}_k \, \forall k$. Therefore, we see that bunching procedure transforms a feasible assignment into new *feasible assignment.* This enables to prove Theorem 1 describing properties of SAOP equilibria.

Theorem 1.[8] *Assume that the cost function $C(.,.)$ is either separable with fixed cost, or differentiable and concave w.r.t. x.*[9] *Then* (**A**): *At any equilibrium $(\bar{x},\bar{t})$, higher nodes in the solution graph $\bar{\bar{G}}(\bar{x},\bar{t})$ bring weakly higher per-package profit than lower nodes (successors) in the sense: $(j \to ...i) \Rightarrow \pi_j(\bar{x},\bar{t}) \geq \pi_i(\bar{x},\bar{t})$ for separable case, or $\bar{\pi}_j(\bar{x},\bar{t}) \geq \bar{\pi}_i(\bar{x},\bar{t})$ for concave case.*

(**B**): *Any optimal solution $(\bar{x},\bar{t})$ can be simplified using a bunching procedure to get another optimal solution $(\hat{x},\hat{t})$, where: (i) different packages bring different profit;*

[7] Firstly, in both definitions, as well as in Theorem 1, variable x_i may be multidimensional. For this case, the derivative in the definition of $\bar{\pi}$ must be replaced by the gradient. Second, the definition should use more rigrous notation $\bar{\pi}_{i\bar{x}}(x,t,)$, because the point $\bar{x}$ of linearization is essential. Hopefully this will not result in any confusion below.

[8] Compare with Guesnerie and Seade (1982), Brito et al. (1990), and Andersson (2005). Our novelties or enforcements are: (1)more general costs, (2)profit monotonicity statement for the *entire* graph, (3)no-splitting claim, and (4)claim that *each* solution can be simplified by bunching (that can be quite important in optimal-taxation setting).

[9] Concavity or non-decreasing returns to scale are essential. The differentiability assumption can be dropped, but at the expense of more cunbersome linearization in the definition of profit per-package.

(ii) higher position in the solution graph relates to strictly higher profit, and (iii) there are no type-splitting and cycles.[10]

PROOF. (**A**). When costs are separable, suppose the contrary to (A) for any couple of adjacent nodes, i.e, suppose a profit-ascending arc $(j \to i) : \bar{\pi}_i = \bar{t}_i - c_i(\bar{x}_i) > \bar{\pi}_j = \bar{t}_j - c(\bar{x}_j)$. Then bunching j to i, as suggested in the above procedure, necessarily increases total profit $(\bar{\pi}(\hat{x}, \hat{t}) > \bar{\pi}(\bar{x}, \bar{t}))$, and it was shown above to generate a *feasible* assignment $(\hat{x}, \hat{t})$. So, the solution $(\bar{x}, \bar{t})$ was not optimal, resulting in a contradiction. This proves (A) for $(j \to i)$.

Similar logic applies under non-separable concave cost for any couple $(j \to i)$ assumed to contradict (A): we can construct an improvement $\bar{\pi}(\hat{x}, \hat{t}) > \bar{\pi}(\bar{x}, \bar{t})$. What remains is to compare artificial linearized profit $\bar{\pi}$ with real profit π. By construction of $\bar{\pi}(.)$, these two functions coincide at the point of linearization $\bar{\pi}(\bar{x}, \bar{t}) = \pi(\bar{x}, \bar{t})$. By concavity, artificial costs are everywhere weakly higher than real costs: $\sum_{i=1}^{n} m_i(c_0 + \bar{c}_i \hat{x}_i) \geq C(m, \hat{x})$, so the real profit is higher than the artificial one, yielding $\pi(\hat{x}, \hat{t}) \geq \bar{\pi}(\hat{x}, \hat{t}) > \bar{\pi}(\bar{x}, \bar{t})$. This again contradicts the optimality of $(\bar{x}, \bar{t})$ and proves (A) for $(j \to i)$.

By induction, the proved claim (A) can be extended from any adjacent couple to any couple (j, i) of graph-comparable nodes.

(**B**). To prove (i), it is sufficient to apply the bunching procedure to eliminate all same-profit nodes. Due to feasibility of bunching procedure, we thus get an *optimal* solution $(\hat{x}, \hat{t})$. So, the statements (ii) and (iii) follow. Q.E.D.

Simply speaking, the theorem states that whenever the principal has separable or/and concave costs (i.e., cost defined per-agent or increasing returns), she may confidently use the SAOP method to design the most profitable package-pricing scheme, without any artificial partitioning, rationing and other special cautions against envy-cycles shown in the examples. Thereby, "stock" tool is also unneeded under concave or separable cost.

[10] Absence of cycle is to be understood here as absence of almost-envy cycles among distinct package-nodes, whereas mutual almost-envy cycle among agents bunched together with the same package is inevitable. No type-splitting means that agents with the same utilities take the same package, even when they are formally represented by different "types" in the model.

5 Conclusions

For discrete consumer-types screening, when cost function reflects capacity constraints or/and standardization, there are examples showing overall distortion of equilibria, resulting from envy-cycles. Thus, under non-separable, non-concave cost, the examples demonstrate that the standard screening model (SAOP) may work inadequately. It is true that this observation raises more questions than it answers. But, we do give some guidelines (e.g., stock variables and partitioning) for incorporating situations with capacity constraints and/or standardization into the screening model, and bring these important issues for future analyses in screening theory. In addition to raising these questions, the realm of situations where SAOP works adequately is extended (by including separable and concave costs), approaching thereby the necessary-and-sufficient conditions for adequacy of using SAOP.

Acknowledgement

The gratitude for financial and professional support is expressed to Economic Education and Research Consortium Inc. (grant R06-0561), financed by Eurasia Foundation, USAID, GDN, World Bank Institute and Government of Sweden. We are grateful also to support by Russian Foundation for Basic Research (rant 07-01-00439a), to seminar participants at CEMI, Moscow, for suggestions, and to our student Denis Anoshin for checking results.

References

Andersson, T.: Profit maximizing nonlinear pricing. Economics Letters 88, 135-139 (2005).

Brito, D.L., Hamilton, J.H., Slutsky, S.M., Stiglitz, J.E. : Pareto efficient tax structures. Oxford Economic Papers 42(1), 61–77 (1990).

Guesnerie, R. and Seade, J.: Nonlinear pricing in a finite economy. Journal of Public Economics 17, 157-179 (1982).

Rochet, J.-C. and Stole, L.: (2007). The economics of multidimensional screening. In: Advances in Economics and Econometrics: Theory and Applications. Cambridge: Cambridge University Press 2007.

Tirole, J.: Theory of industrial organization, Cambridge, MA, MIT Press 1994

Figures

Figures should be incorporated in the text as marked.

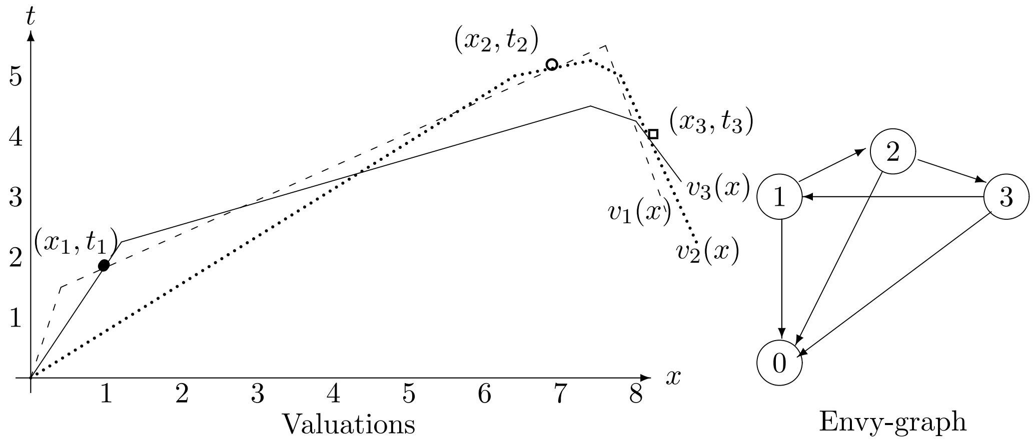

Figure 1: Generic inefficient cycle under convex cost.

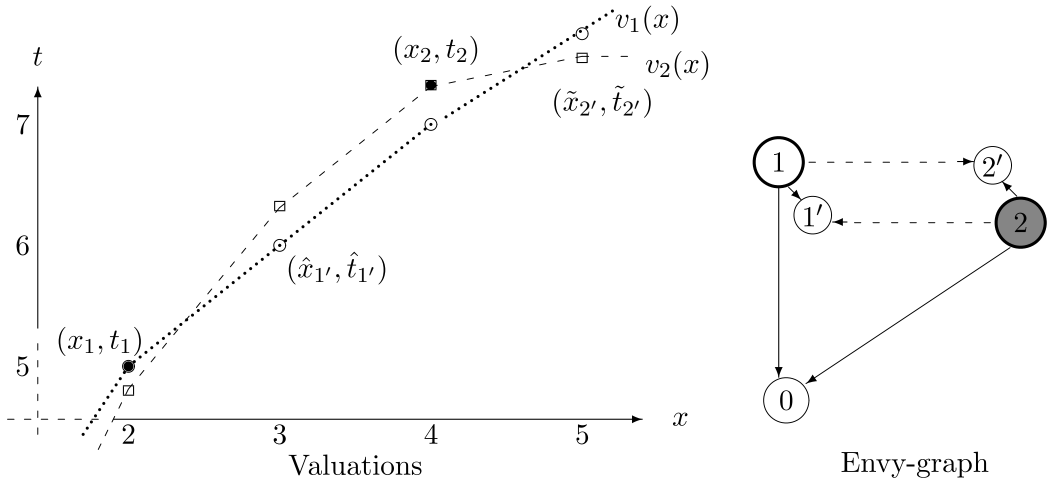

Figure 2: Generic inefficient shadow-cycle under discrete cost.

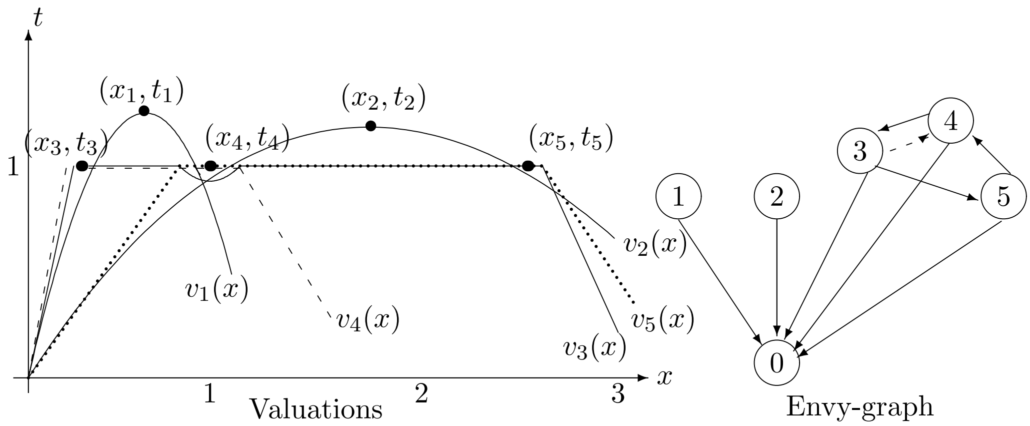

Figure 3: Non-reducible cycle for locally-similar agents and convex cost.